CIEEM

Metadata Standards

Professional Guidance Series
PGS9

www.cieem.net

Introduction

Metadata are additional sets of information supplied to facilitate the understanding and use of data by describing the origin, content, purpose, format, location and limitations of a given data set.

Ecological data are collected for a variety of reasons and using a wide range of methods. Data may be passed to clients, colleagues, and other interested bodies, or deposited in an open-access archive. The data may or may not be incorporated into a formal report, but common to all situations is the need for a comprehensive set of metadata to enable potential users to assess whether the data are 'fit for purpose'.

End-users of an ecological data set may not have been involved in the planning, execution or reporting stages of the original survey. Surveyors, authors of reports, and managers of data archives therefore have a responsibility to ensure that metadata are freely and fully available. It cannot be assumed that end-users of a data set will know which data collection methods have been used or understand potential weaknesses in the data.

A metadata standard has two elements defining:

- Guidance on the set of information to accompany a dataset and necessary for full and appropriate use of the data
- Principles of good practice to be followed in the management of data and the accompanying metadata.

This issue in the Professional Guidance Series outlines general principles of relevance to members of CIEEM in professional practice. It is recommended that metadata management is incorporated into members' in-house quality assurance procedures in reporting, sharing and archiving of ecological data.

Metadata Defined

Metadata are defined as '*the information necessary to understand and effectively use data, including documentation of the data set content, contexts, quality, structure and accessibility*' (Michener 2000). A simple metadata set would answer the following questions: What are the data? Why, when, where and how were data collected, and by whom? However, comprehensive metadata should cover much more; enabling the end-user to assess the reliability of the information provided, and its suitability for a given application.

A data set without metadata is greatly degraded, potentially misleading, and represents a wasted investment in the initial survey. A well executed survey may be undermined by inadequacies in subsequent reporting. Adopting a metadata standard for reporting, sharing and archiving ecological data is therefore good practice and complies with the CIEEM Code of Professional Conduct; particularly with regard to the need to identify the limitations of survey data.

Defining a minimum set of metadata is only part of the approach towards adoption of a metadata standard. Members are also recommended to adopt an explicit policy towards documentation, storage and dissemination of metadata.

Metadata Management Principles

Principles contributing to efficient data management include ensuring that metadata are:

- Recorded fully by the originator of the data
- Stored with the 'raw' data in a portable format that can be transmitted with the data
- Documented whenever derived summary statistics, or the original data, are incorporated into written reports
- Recognised as being an integral part of the data set
- Shared freely wherever possible.

Metadata allow an assessment of the reliability of data *per se* and conclusions based upon these data. It follows that:

- The field surveyor has responsibilities in the documentation of metadata and passing it on accordingly.
- Metadata should be documented fully in survey reports produced for clients, but should also be stored with the raw data for subsequent use and dissemination.

- Where data are shared with outside bodies and individuals the metadata should be transmitted with the dataset.
- Organisations receiving data from third parties, or contracting surveyors, should check that all data are accompanied by relevant metadata.
- It may be helpful to specify metadata reporting standards in contracts with surveyors and data suppliers.
- Particular care should be taken to ensure that key elements of the metadata are not omitted when editing reports written by others, or incorporating elements of third-party reports into other documents.

Many of these concerns could be addressed by using a metadata pro-forma or standardised record sheet that is stored with the data. An outline metadata pro-forma is shown in Appendix 1 (see below). The checklist addresses the following aspects of the data:

- Background to data collection
- Content of the data set
- Data collection methods
- Quality assurance procedures operating during and after the field survey
- Data archiving and access.

Members may wish to build on this preliminary checklist to suit specific methods, habitats and taxa.

A Minimum Metadata Set

The specific metadata required to enable full interpretation of ecological data will vary greatly between projects according to the taxa and spatiotemporal scales involved. Moreover, it is important to recognise that Appendix 1 is not intended as an outline of the content of a survey report; rather it is a guide to the ideal metadata to be included therein, and stored, and disseminated with a data set. For further guidance on report writing, see CIEEM's Guidance on 'Ecological Report Writing'.

Metadata Standards, Survey Design and Data Quality

A metadata standard can be used as a vehicle to improve standards in data collection by encouraging a focus on data quality issues. Improving the quality of ecological survey data requires information on current levels of precision and accuracy.

Some aspects of data quality can be addressed after completion of field work (e.g. assessing adequacy of sample sizes); while other aspects (e.g. quantification of observer precision and accuracy) are best assessed during the survey itself. It follows that early consideration of the requirements for the reporting of metadata will be beneficial in focussing attention on quality assurance as a necessary element of field survey design. This is a neglected area of reporting in ecology and environmental management. Adoption of an ecological metadata standard with a specific emphasis on quality assurance will pay dividends in raising awareness of the potential problems that can then be addressed proactively.

Links to National Metadata Standards

Appendix 1 has been drawn up as a checklist that could be used to review the adequacy of metadata accompanying a data set or survey report. It is based broadly on the content of the UK GEMINI (Geo-spatial Metadata Interoperability Initiative) Metadata Standard, although with emphasis given to aspects most likely to be of significance in ecological reporting. The UK GEMINI Standard was produced by the Association for Geographic Information (AGI) in collaboration with the UK Cabinet Office e-Government Unit (with additional representation from national and local government, and the GIS community). It is the UK standard used by government bodies and is the framework used for submission of metadata to the National Biodiversity Network (NBN) (www.nbn.org.uk).

Further Information

Those wishing to read more about metadata in an ecological context may find the following publications of use:

Association for Geographic Information (2010). UK GEMINI: Specification for discovery metadata for geospatial data resources, version 2.1, August 2010.

Chartered Institute of Ecology and Environmental Management (2017) *Guidelines for Ecological Report Writing.* Available at: https://cieem.net/resource/guidelines-for-ecological-report-writing/

Michener WK (2000). Metadata in: Michener WK and Brunt JW (eds) *Ecological data. Design, management and processing*. Blackwell Science, Oxford.

Michener WK, Brunt JW, Helly JJ, Kirchener TB and Stafford SG (1997). Nongeospatial metadata for the ecological sciences. *Ecological Applications* 7: 330- 342.

Acknowledgements

This guidance note was written by Dr Andrew Cherrill with some assistance from the Professional Standards Committee.

Appendix 1

This contains a provisional checklist of metadata (see overleaf) to accompany an ecological data set. It has been drawn up as a checklist that could be used to review the adequacy of metadata accompanying a data set or survey report. It is based broadly on the content of the UK GEMINI (Geo-spatial Metadata Interoperability Initiative) Metadata Standard, although with emphasis given to aspects most likely to be of significance in ecological reporting.

Topic	Detail of metadata to accompany data/survey report	Included Yes/No
Title	• Title of the dataset (typically including taxa, data type, date(s) and geographical extent of information)	
Subject	• A shortlist of keywords that describe the dataset	
Language	• Language used in the dataset if not English (e.g. Welsh, Gaelic)	
Abstract	• A clear and concise statement that enables the reader to understand the content of the dataset • Circumstances leading to data collection • Explicit aims of the survey(s) yielding the data • Brief overview of dataset content	
Ownership	• Ownership and copyright of data • Contact name and address for further information and support to facilitate access to, and use of, data	
Survey date	• Date(s) of data collection (DD/MM/YY) as single date(s) or range • Time of day of start / finish and weather conditions (if relevant and not part of methods e.g. butterfly transects) • Frequency of update if survey is ongoing	
Location	• Name of the site(s), county and country • Longitudinal and latitudinal bounds of survey extent if sample points are widely scattered • Site description (major habitats and altitudinal range if this information is not part of data set) • Site ownership and access (where possible)	
Survey personnel	• Name(s) of field surveyors and organisation(s) • Experience of surveyors in the methods used • Qualifications, professional memberships and licences held (as relevant)	
Field methods	• Description of methods sufficient to permit repetition and assess adequacy of sampling design • Detailed site map showing location of sample points, transects and/or routes followed during walksurveys • Reference to published manuals and project specific modifications to standard methods • Brief justification of the approach taken	
Data descriptors	• Data presentation type (e.g. digital map or spreadsheet) • Definition of the data variables • Units of measurement • Precision of measurements (e.g. to one decimal point) • Spatial resolution (e.g. minimum mappable unit) and for GIS records spatial representation (e.g. vector or grid) • Spatial reference system used in data (e.g. British National Grid Reference or Post Code) • Detailed key to data labels where abbreviations are used in dataset (e.g. in rows and columns of spreadsheet) • Full reference to species identification keys and habitat classifications used • Taxonomic authority cited for scientific binomials	
Data processing	• Name(s) of dataset compiler(s) and organisation(s) • Date of dataset compilation (DD/MM/YY) • Description of any processing of data after collection (e.g. error correction) and name/version of any software utilised (e.g. GIS package) • Justification for processing decisions	
Quality assurance	• Methods used to determine accuracy of observations, and species identification during survey (e.g. re-sampling) • Post-survey assessments of data reliability (e.g. adequacy of sample size) • Conclusions drawn from these QA procedures; including, for example, quantitative estimates of data accuracy.	
Data archiving and usage	• Location of the 'raw' data • Name, date and version of computer software with which data have been saved and file type (e.g. .doc,.bmp) • Restrictions on access to the data • Constraints on use of the data • Record of changes to data and metadata with name and organisation of editor(s) • Details of reports/publications describing or using the data • Additional sources of information (e.g. publications or websites that describe background to the project or data)	

Raising the profile of professional ecological and environmental management and promoting the highest standards of practice for the benefit of nature and society.

CIEEM
Chartered Institute of Ecology and Environmental Management

43 Southgate Street
Winchester, Hampshire SO23 9EH

t: 01962 868626
e: enquiries@cieem.net
www.cieem.net